Electronic Version
Stylesheet Version v1.1.1

Description

[METHOD OF FABRICATING A FLASH MEMORY]

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the priority benefit of Taiwan application serial no. 93103004, filed February 10, 2004.

BACKGROUND OF INVENTION

[0002] Field of the Invention

[0003] The present invention relates to a method of fabricating a memory device. More particularly, the present invention relates to a method of fabricating a flash memory and floating gate.

[0004] Description of Related Art

[0005] Flash memory is a type of electrically erasable programmable read-only memory (EEPROM). Flash memory is a memory device that allows multiple data writing, reading and erasing operations. The stored data will be retained even after power to the device is removed. With these ad-

vantages, it has been broadly applied in personal computer and electronic equipment. In addition, the flash memory is also a type of high-speed non-volatile memory (NVM) that occupies very little space and consumes very little power. Moreover, erasing is carried out in a block-by-block fashion so that the operating speed is higher than most conventional memory devices.

[0006] A typical flash memory device has a floating gate and a control gate formed by doped polysilicon. The control gate is set up directly above the floating gate with an inter-gate dielectric layer separating the two. Furthermore, a tunneling oxide layer is also set between the floating gate and the underlying substrate (the so-called stacked gate flash memory). To operate the flash memory, a positive or negative voltage is applied to the control gate so that electric charges can be injected into or released from the floating gate resulting in the storage or erasure of data.

[0007] Figs. 1A through 1C are schematic cross-sectional views showing some of the steps for fabricating a conventional flash memory device. First, as shown in Fig. 1A, a substrate 100 having a plurality of device isolation structures 102 thereon for defining active regions 104 and a tunnel-

ing dielectric layer on the active regions 104 is provided. A conductive layer 108 is formed over the substrate 100 to cover the device isolation structures 102 and the tunneling dielectric layer 106. Thereafter, a planarization operation is carried out to remove a portion of the conductive layer 108 and smooth out the top surface of the conductive layer 108.

[0008] As shown in Fig. 1B, a patterned photoresist layer 109 is formed over the conductive layer 108. The patterned photoresist layer 109 exposes a portion of the conductive layer 108 on the device isolation structure 102. Thereafter, using the patterned photoresist layer 109 as a mask, a portion of the conductive layer 108 is removed to form a plurality of trenches 107 in the conductive layer 108 above the device isolation structures 102. The conductive layer 108 retained after forming the trenches 107 becomes the floating gate 110.

[0009] After removing the patterned photoresist layer 109, an inter-gate dielectric layer 112 is formed over the substrate 100 to cover the floating gate 110 as shown in Fig. 1C. Finally, a control gate 114 is formed over the inter-gate dielectric layer 112.

[0010] In the aforementioned fabrication process, the floating

gate 110 is formed using photolithographic and etching processes. However, photolithographic and etching processes involve steps such as de-moisturize heating, coating, photoresist deposition, soft baking, photo-exposure, post photo-exposure baking, chemical development, hard baking and etching. Hence, the process not only is time consuming but also incurs additional production cost.

[0011] In addition, the aforementioned process utilizes a chemical-mechanical polishing (CMP) operation to planarize the conductive layer 108. Without a reference polishing stop layer, the thickness of conductive layer 108 retained after each chemical-mechanical polishing operation will be different. In other words, there is no control over to the thickness of the floating gate 110.

[0012] On the other hand, a larger gate-coupling ratio (GCR) between the floating gate and the control gate requires a lower operating voltage. The methods of increasing the gate-coupling ratio include increasing the capacitance of the inter-gate dielectric layer or reducing the capacitance of the tunneling oxide layer. One method of increasing the capacitance of the inter-gate dielectric layer is to enlarge the included area between the control gate and the floating gate. Thus, minimizing the size of the trenches

107 is able to increase the included area between the floating gate and the control gate and thus increase the gate-coupling ratio between them. However, when the conductive layer 108 is patterned, size of the trenches 107 is constrained by the photolithographic and etching processes. In other words, it is difficult to decrease the size of each trench 107 further. In the absence of any other method for increasing the included area between the control gate and the floating gate, improving the performance of the memory device is difficult.

SUMMARY OF INVENTION

[0013] Accordingly, the present invention is directed to a method of fabricating a flash memory capable of controlling the thickness of a floating gate inside the flash memory and increasing the gate-coupling ratio between the floating gate and a control gate for a higher device performance.

[0014] The present invention is also directed to a method of fabricating a floating gate such that there is no need to fabricate the mask for forming the floating gate. In other words, one photolithographic and etching process can be effectively avoided so that the fabricating process is more simplified.

[0015] According to an embodiment of the present invention, a

method of fabricating a flash memory is provided. First, a substrate with a tunneling dielectric layer, a first conductive layer, a pad oxide layer and a patterned mask layer sequentially formed thereon is provided. Thereafter, using the patterned mask layer as a mask, a portion of the pad oxide layer, the first conductive layer, the tunneling dielectric layer and the substrate are removed to form a plurality of first trenches in the substrate. Insulating material is deposited into the first trenches to form a plurality of device isolation structures. A portion of each device isolation structure is removed to form a plurality of second trenches such that the top section of each retained device isolation structure lies between the tunneling dielectric layer and the patterned mask layer. A dielectric layer is formed over the substrate to cover the patterned mask layer and the surface of the second trenches. Material is deposited into various second trenches to form a sacrificial layer. The sacrificial layer and the dielectric layer are formed by different materials each having a different etching selectivity. Using the sacrificial layer as a self-aligned mask, a portion of the dielectric layer is removed. The patterned mask layer is removed to expose the pad oxide layer and then the pad oxide layer is re-

moved to expose the first conductive layer. Thereafter, a second conductive layer is formed over the substrate. A portion of the second conductive layer is removed to expose the top section of the sacrificial layer. The second conductive layer and the first conductive layer together constitute a floating gate. The method of removing a portion of the second conductive layer to expose the top section of the sacrificial layer includes performing a chemical-mechanical polishing operation. Furthermore, the second conductive layer and the sacrificial layer are formed by different materials each having a different etching selectivity. Thereafter, the sacrificial layer is removed. An inter-gate dielectric layer is formed over the substrate to cover the floating gate. A control gate is formed over the inter-gate dielectric layer. Finally, a source region and a drain region are formed in the substrate on each side of the control gate.

[0016] In the process of forming the floating gate, the second trenches are formed over the device isolation structures before sequentially depositing the dielectric material and sacrificial material into the second trenches to form a stack structure. Thereafter, the stack structure is used to fabricate the floating gate. Hence, the present invention

eliminates a mask for fabricating the floating gate. In other words, one photolithographic and etching process can be effectively avoided and hence the overall fabrication cost can be reduced.

[0017] Because the thickness of the floating gate correspond to the total height of the dielectric layer and the sacrificial layer, the thickness of the floating gate is determined by the total height of the dielectric layer and the sacrificial layer. Thus, the thickness of the floating gate can be precisely controlled.

[0018] In addition, the size of the second trenches can be reduced by forming a thicker dielectric layer. Hence, a floating gate with a larger size can be produced. With a larger floating gate, the included area between the control gate and the floating gate is increased so that a higher gate-coupling ratio is obtained.

[0019] The present invention also provides an alternative method of fabricating a flash memory. First, a substrate with a plurality of device isolation structures for defining active regions and a tunneling dielectric layer and a patterned mask layer sequentially formed over the substrate within the active regions is provided. Thereafter, a portion of each device isolation structure is removed to form a plu-

rality of trenches. The top section of each retained device isolation structure lies between the tunneling dielectric layer and the patterned mask layer. A dielectric layer is formed over the substrate to cover the patterned mask layer and the surface of the trenches. Sacrificial material is deposited into the trenches to form a sacrificial layer. The sacrificial layer and the dielectric layer are formed by different materials each having a different etching selectivity. Using the sacrificial layer as a self-aligned mask, a portion of the dielectric layer is removed. Thereafter, the patterned mask layer is removed to expose the tunneling dielectric layer. A conductive layer is formed over the substrate. Afterwards, a portion of the conductive layer is removed to expose the top section of the sacrificial layer. The method of removing a portion of the conductive layer to expose the top section of the sacrificial layer includes performing a chemical-mechanical polishing operation. Furthermore, the conductive layer and the sacrificial layer are formed by different materials each having a different etching selectivity. Finally, the sacrificial layer is removed.

[0020] In the process of forming the floating gate, the trenches are formed over the device isolation structures before sequentially depositing the dielectric material and sacrificial

material into the trenches to form a stack structure. Thereafter, the stack structure is used to fabricate the floating gate. Hence, the present invention eliminates the need to fabricate a mask for fabricating the floating gate. In other words, one photolithographic and etching process can be effectively avoided and hence the overall fabrication cost can be reduced.

[0021] Because the thickness of the floating gate correspond to the total height of the dielectric layer and the sacrificial layer, the thickness of the floating gate is determined by the total height of the dielectric layer and the sacrificial layer. Thus, the thickness of the floating gate can be precisely controlled.

[0022] It is to be understood that both the foregoing general description and the following detailed description are exemplary, and are intended to provide further explanation of the invention as claimed.

BRIEF DESCRIPTION OF DRAWINGS

[0023] The accompanying drawings are included to provide a further understanding of the invention, and are incorporated in and constitute a part of this specification. The drawings illustrate embodiments of the invention and, together with the description, serve to explain the principles

of the invention.

[0024] Figs. 1A through 1C are schematic cross-sectional views showing some of the steps of fabricating a conventional flash memory device.

[0025] Figs. 2A through 2F are schematic cross-sectional views showing the steps of fabricating a flash memory according to one embodiment of the present invention.

DETAILED DESCRIPTION

[0026] Reference will now be made in detail to the present preferred embodiments of the invention, examples of which are illustrated in the accompanying drawings. Wherever possible, the same reference numbers are used in the drawings and the description to refer to the same or like parts.

[0027] Figs. 2A through 2F are schematic cross-sectional views showing the steps for fabricating a flash memory according to one embodiment of the present invention. As shown in Fig. 2A, a substrate 200 such as a silicon substrate is provided. Thereafter, a tunneling dielectric layer 206, a conductive layer 208, a pad oxide layer 209 and a patterned mask layer 210 are sequentially formed over the substrate 200. The patterned mask layer 210 has openings 202 that expose areas for forming a device isolation

structure.

[0028] The tunneling dielectric layer 206 is silicon oxide layer having a thickness between about 70Å to 90Å formed, for example, by performing a thermal oxidation process. The conductive layer 208 is a doped polysilicon layer formed, for example, by performing a chemical vapor deposition process to form an undoped polysilicon layer (not shown) and then implanting ions into the undoped layer to form a doped polysilicon layer having a thickness between about 500Å to 1000Å. The pad oxide layer 209 is a silicon oxide layer having a thickness between about 15Å to 50Å formed, for example, by performing a thermal oxidation process. Furthermore, the patterned mask layer 210 is formed by a material having an etching selectivity that differs from the pad oxide layer 209, the conductive layer 208, the tunneling dielectric layer 206 and the substrate 200. The patterned mask layer 210 is a silicon nitride layer having a thickness between about 1500Å to 2000Å, for example. The patterned mask layer 210 is formed, for example, by performing photolithographic and etching processes.

[0029] As shown in Fig. 2B, a portion of the pad oxide layer 209, the conductive layer 208, the tunneling dielectric layer

206 are removed using the patterned mask layer 210 as an etching mask to form a plurality of trenches 212. Ultimately, a tunneling dielectric layer 206a, a conductive layer 208a and a pad oxide layer 209a remain on top of the substrate 200. The trenches 212 have a depth of, for example, between about 3000Å to 4000Å.

[0030] Thereafter, an insulating material is deposited into the trenches 212 to form a plurality of device isolation structure 214 for defining an active region 204. The device isolation structure 214 is formed, for example, by performing a high-density plasma chemical vapor deposition (HDP-CVD) process to form a layer of insulation material (not shown) and then performing a chemical-mechanical polishing (CMP) operation to remove material outside the trenches.

[0031] It should be noted that, in this embodiment, the tunneling dielectric layer 206 is formed before forming the device isolation structures 214. This can prevent the formation of bird's beak in the neighborhood of the device isolation structure due to a subsequent thermal process if the device isolation structure 214 is formed first.

[0032] As shown in Fig. 2C, a portion of the insulation material in each device isolation structures 214 is removed to form a

plurality of trenches 215. A top section of the remaining device isolation structure 214a lies between the tunneling dielectric layer 206a and the patterned mask layer 210. The method of removing a portion of the insulation material from the device isolation structures 214 to form the trenches 215 includes a dry etching process.

[0033] Thereafter, a dielectric layer 216 is formed over the substrate 200 to cover the patterned mask layer 210 and the surface of the trenches 215. The dielectric layer 216 is formed by a material having an etching selectivity that differs from the material for forming a conductive layer in a subsequent process. The dielectric layer 216 is a silicon nitride layer having a thickness between about 200Å to 1000Å formed, for example, by performing a chemical vapor deposition process. In this embodiment, both the dielectric layer 216 and the patterned mask layer 210 are formed by an identical material.

[0034] Sacrificial material is deposited into each trench 215 to form a sacrificial layer 218. The sacrificial layer 218 is formed by a material having an etching selectivity that differs from the material for forming a conductive layer in a subsequent process. The sacrificial layer 218 is a silicon oxide layer formed, for example, by depositing a layer of

sacrificial material (not shown) and then performing a chemical-mechanical polishing operation or a back-etching process to remove sacrificial material lying outside the trenches 215. In another preferred embodiment, the sacrificial layers 218 are formed, for example, by spin-coating a layer of spin-on glass (SOG) over the substrate 200 to form a sacrificial layer (not shown) and then etching back the excess sacrificial material outside the trenches 215.

[0035] As shown in Fig. 2D, using the sacrificial layers 218 as a self-aligned mask, a portion of the dielectric layer 216 is removed. Since the sacrificial layers 218 and the dielectric layer 216 are fabricated from materials having a different etching selectivity, most of the dielectric layer 216 is removed except the dielectric layer 216a underneath the sacrificial layers 218. The dielectric layer 216a and the sacrificial layer 218 together form a sacrificial stacked layer 217. Because the dielectric layer 216 and the patterned mask layer 210 are formed by the same material (for example, silicon nitride) in this embodiment, the process of removing a portion of the dielectric layer 216 also removes the patterned mask layer 210.

[0036] Thereafter, the pad oxide layer 209a is removed to expose

the conductive layer 208a. The pad oxide layer 209a is removed, for example, by wet etching using hydrofluoric acid solution as the etchant. A conductive layer 220 is formed over the substrate 200. With the conductive layer 208a already formed underneath, the conductive layer 220 is easier to form on top. In addition, the conductive layer 220 is formed by doped polysilicon, for example. The doped polysilicon layer is formed, for example, by performing a chemical vapor deposition process to form an undoped polysilicon layer (not shown) and then implanting ions into the undoped polysilicon layer.

[0037] As shown in Fig. 2E, a portion of the conductive layer 220 is removed to expose the top section of the sacrificial layer 218 so that the retained conductive layer 220a and the conductive layer 208a together constitute a floating gate 221. The method of removing a portion of the conductive layer 220 to expose the top section of the sacrificial layer 218 includes performing a chemical-mechanical polishing operation using the sacrificial layer 218 as a polishing stop layer. Hence, the retained conductive layer 220a has a thickness related to the total height of the sacrificial stacked layer 217. In other words, a better control of the thickness of the floating gate 221 is obtained.

[0038] It should be noted that the thickness of the dielectric layer 216 on the sidewalls of the trenches 215 in Fig. 2C directly affects the size of the conductive layer 220a. That is, it also affects the overlapping area between the floating gate 221 and the control gate (not shown). Consequently, in the aforementioned step, a thicker dielectric layer 216 can be used to reduce the width of the trench 215 so that the distance between neighboring conductive layers 220a can be reduced. For example, in Fig. 2C, if the original width W1 of the trench 215 is 2000Å and the width W2 of the patterned mask layer 210 between two trenches 215 is 1500Å, the width W3 of the trench 215 would be 1000Å after depositing a dielectric layer 216 with a thickness of about 500Å. Hence, the conductive layer 220a originally having a maximum width of about 1500Å (width W2 of the patterned mask layer 210) can have a wider width W4 of about 2500Å as shown in Fig. 2E. In other words, electrical performance of the memory device can be increased by forming a thicker dielectric layer 216 to increase the overlapping area between the floating gate 221 and the control gate.

[0039] As shown in Fig. 2F, the sacrificial layers 218 are removed. The sacrificial layers 218 are removed, for exam-

ple, by wet etching using hydrofluoric acid solution as the etchant. It should be noted that, in this embodiment, the trenches 215 are formed before the sacrificial stacked layer 217 that includes the dielectric layer 216 and the sacrificial layer 218 being formed. Then, the floating gate 221 is formed utilizing the sacrificial stacked layer 217 as the etching stop layer. Consequently, one photolithographic process is omitted and the production cost is reduced.

[0040] Thereafter, an inter-gate dielectric layer 222 is formed over the substrate 200 to cover the dielectric layer 216a and the floating gate 221. The inter-gate dielectric layer 222 is an oxide/nitride/ oxide composite layer, for example. The inter-gate dielectric layer 222 is formed, for example, by performing a thermal oxidation process to form a silicon oxide layer over the substrate 200 and then performing a chemical vapor deposition process to form a silicon nitride layer and another silicon oxide layer over the first silicon oxide layer. The oxide/nitride/oxide composite layer has a first oxide layer with a thickness between 40Å to 50Å, a silicon nitride layer with a thickness between 45Å to 70Å and a second silicon oxide between 50Å to 70Å. Obviously, the inter-gate dielectric layer 222

can be an oxide/nitride composite layer too.

[0041] A control gate 224 is formed over the inter-gate dielectric layer 222. The control gate 224 is a doped polysilicon formed, for example by performing a chemical vapor deposition process to form a layer of undoped polysilicon (not shown) and implanting ions into the undoped polysilicon layer. Thereafter, a source region (not shown) and a drain region (not shown) are formed in the substrate on each side of the control gate 224. The source region and the drain region are formed, for example, by implanting impurities into the substrate 200 on each side of the control gate 224. Since subsequent fabrication processes should be familiar to those skilled in the techniques, detailed description is omitted here.

[0042] Aside from the aforementioned embodiment of the present invention, it should be noted that there is another embodiment. After removing the pad oxide layer 208a in Fig. 2D, the conductive layer 208a is removed before carrying out the step for forming the conductive layer 220 and the processes as shown in Figs. 2E and 2F. Hence, the completed flash memory has a floating agate 221 including just the conductive layer 220a. Furthermore, in another preferred embodiment, a substrate 200 with only a

tunneling dielectric layer 206 and a patterned mask layer 210 thereon is provided in Fig. 2A. Thus, the floating gate 221 of the flash memory also includes a single conductive layer 220a only. In yet another preferred embodiment, after removing the sacrificial layers 218 in Fig. 2F, further includes removing the dielectric layer 216a before carrying out the steps for forming the inter-gate dielectric layer 222 and the control gate 224.

[0043] In summary, major advantages of the present invention includes:

[0044] 1. Trenches are formed over the device isolation structures before depositing dielectric material and sacrificial material into them to form a stack structure. Thereafter, the stack structure is used to fabricate the floating gate. Hence, the present invention eliminates the need to fabricate a mask for fabricating the floating gate. In other words, one photolithographic and etching process can be effectively avoided and hence the overall fabrication cost can be reduced.

[0045] 2. Because the thickness of the floating gate correspond to the total height of the dielectric layer and the sacrificial layer, the thickness of the floating gate is determined by the total height of the dielectric layer and the sacrificial

layer. Thus, the thickness of the floating gate can be precisely controlled.

[0046] 3. With the size of the second trenches reduced by forming a thicker dielectric layer, a floating gate with a larger size can be produced. With a larger floating gate, the included area between the control gate and the floating gate is increased so that a higher gate-coupling ratio and hence a better electrical performance of the device is obtained.

[0047] 4. The tunneling dielectric layer is formed before carrying out various steps for fabricating the device isolation structures. This can prevent the formation of bird"s beak in the neighborhood of the device isolation structure due to a subsequent thermal process when the device isolation structure is formed first. Ultimately, the electrical performance of the memory device is improved.

[0048] It will be apparent to those skilled in the art that various modifications and variations can be made to the structure of the present invention without departing from the scope or spirit of the invention. In view of the foregoing, it is intended that the present invention cover modifications and variations of this invention provided they fall within the scope of the following claims and their equivalents.